Embedding Suitability Adaptive Cover Selection for Image Steganography

Junying Yuan[1], Haishan Chen[2]

[1]Department of ECSE, Nanfang College of Sun Yat-Sen University, Guangdong Province, China
[2]School of SIST Sun Yat-Sen University, Guangdong Province, China
cihisa@126.com, hysonyz@126.com

Abstract - There have been a number of state-of-art spatial image steganography algorithms which could hide data with a high degree of undetectability against steganalysts by embedding information adaptively into noisy regions. However, there is still a gap from real application when selecting a suitable cover image for a given payload. This paper tries to minimize this gap by presenting a novel cover selection scheme which measures image embedding suitability under user constraints. Experiments show that the acquired empirical security exceeds the expectation.

Index Terms - Steganography, Cover Selection

1. Introduction

Steganography is the science of secret communication by hiding information in different media such as images, audios, videos etc without making visual changes. Content adaptive image steganography embeds data in noisy image regions wherein the pixels are hard to model so as to reduce image statistical distortion. The most successful approaches formulate data hiding as a source coding with fidelity constraint problem and the problem is resolved by a very general class of distortion functions [1, 2]. Distortion functions are heuristically designed to measure the statistical distortion on changing cover elements.

In spatial image steganography, much effort has been put on designing distortion functions to improve empirical security. Edge adaptive steganography [3] embeds information in the sharper edges according to the embedding rate while avoiding impacting the smooth regions. Highly Undetectable SteGO (HUGO) [4] hides information bits in the noisy pixels that are hard to model in SPAM features. Wavelet Obtained Weights (WOW) and UNIversal WAvelet Relative Distortion (UNIWARD) restrict embedding to regions with rich textures by employing Wavelet-based Directional Filter Bank 'WDFB-D' to compute the directional residuals which measures the embedding suitability.

While most of those steganography algorithms are empirically verified effective in the laboratory, several problems still remains open before putting them into real world application [7]. One of the problems is how to select a suitable cover under constraints of given security and payload. In [8], a cover selection scheme is proposed to select a suitable cover in a blind way from a set of covers by comparing the obtained stego against the cover with distortion measurements of number of modifications, MSE and/or prediction error. Such a method is ineffective as it has no knowledge of what kind of covers can meet the requirement. As far as the author knows, there is no other cover selection scheme available.

This paper tries to make spatial image cover selection more realistic by proposing a novel cover selection scheme using prior knowledge of image embedding suitability which is usually the inverse of embedding distortion under constraints of empirical security and payload.

2. Preliminaries

The following notation conventions are used in this paper for better readability. Capital and lower-case boldface symbols are used solely for matrices and vectors. The symbols $\mathrm{X} = \left(\chi_{ij}\right), \mathrm{Y} = \left(Y_{ij}\right) \in \left\{0...255\right\}^{n1\times n2}$ stand for the matrices of a gray-scale cover image and the corresponding stego image with $\mathrm{n1} \times \mathrm{n2}$ pixels.

A. *Empirical Security*

All experiments are performed on Boss V1.0 [4] which contains 10000 gray-scale images of 512×512 pixels. Edge adaptive LSBM [3] and HUGO [4] are used as steganography algorithms for verification. Empirical security is evaluated in the ensemble's "Out-Of-Bag" (OOB) using SPAM [9] model consisting 686 dimensional features which is designed for spatial images.

B. *Embedding Suitability*

We use the term "embedding suitability" to measure the suitability of a cover pixel to hide information. It is usually defined as the inverse of embedding distortion [6]. The larger the embedding suitability is, the better the pixel is to hide information. In image data hiding, the hard-to-model (or complex) pixels are heuristically suitable for information hiding [3][4][5][6]. So even there is no standard definition of "embedding suitability", we can measure it by the complexity of a pixel in its local region.

In this paper, we use the directional filter residuals to quantify embedding suitability. Ref. [6] proved experimentally that the empirical security is better when using directional filter bank 'WDFB-D' as embedding suitability measurement comparing to that of directionless filter KB. So filter residuals output from WDFB-D are used in this paper to measure a pixel's embedding suitability. WDFB-D directional residuals $\Re_{ij}^{k}$ have to be combined together as a pixel's single embedding suitability ξ_{ij}, as in (1). Then the sum of all pixels' embedding suitability is used to evaluate an image's global embedding suitability. We restrict the sum up of embedding suitability to additive form for simplicity as the way summing up distortion. Image embedding suitability is made scale-independent by defining it as a per pixel average quantity $\Gamma(\mathrm{X})$, as in (2).

$$\xi_{ij} = \sum_{k=1}^{K} \Re_{ij}^{k}, K = 3 \qquad (1)$$

$$\Gamma(\mathrm{X}) = \frac{1}{n1 \times n2} \sum_{i=1}^{n1} \sum_{j=1}^{n2} \xi_{ij} \qquad (2)$$

An experiment is performed to verify if empirical security changes following image embedding suitability. Firstly, $\Gamma(\mathrm{X})$ is calculated for all figures in Boss V1.0 and arrange images in the order of $\Gamma(\mathrm{X})$. Then the ordered sequence of images is cut into groups of 2000 images with 1800 images overlapping following image embedding suitability order, and the average embedding suitability of each image group is used to represent the group's embedding suitability to improve cover selection efficiency. Finally empirical security OOB is obtained from ensemble classifier on each image group with various payloads.

The normalized image group embedding suitability and empirical security of HUGO with normalized payload (two times of payload) is illustrated in a Fig. 1. We can see clearly from Fig. 1 which cover group meets the given payload and empirical security for steganography HUGO. The cyan line in Fig. 1 illustrates the normalized $\Gamma(\mathrm{X})$ of all image groups in descending order of embedding suitability. In the former part where $\Gamma(\mathrm{X})$ is high enough, OOB degrades as embedding suitability drops. However, in the latter part where $\Gamma(\mathrm{X})$ is too low, OOB improves as embedding suitability drops. One possibility is that SPAM features are not good enough to measure low complexity images. This is a minor problem as this paper focuses on cover selection scheme.

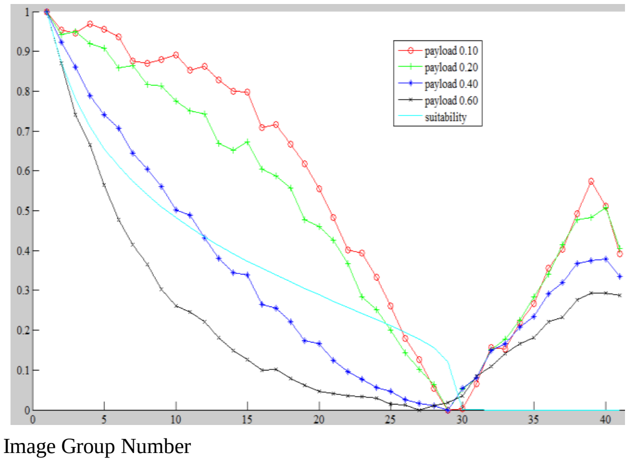

Fig. 1: Embedding suitability of image groups in descending order and the OOB of HUGO for each image group with normalized payload (two times of payload) of 0.10, 0.20, 0.40 and 0.60.

A further experiment for steganography algorithm edge adaptive LSBM gives similar results. The experimental result proves that given the user constraints of <steganography algorithm, payload, minimum OOB>, suitable cover can be found following the prior knowledge as shown in Fig. 1. In another word, if an image statistic database as the prior knowledge is available, we can select suitable image covers directly from the database, and this is where the proposed cover selection scheme comes from.

3. Cover Selection Scheme Design

This section first details the proposed cover selection scheme based on prior knowledge of image embedding suitability, steganography algorithms, payload and empirical security, then illustrates on the prior knowledge constructions and finally makes an analysis of the experimental results.

A. *Proposed Cover Selection Scheme*

We propose a cover selecting scheme based on image embedding suitability as what Fig. 2 illustrates. The scheme requires input of user requirement, including steganography algorithm, payload, and expected worst empirical security OOB. The covers are selected by searching the image statistic database as the prior knowledge. The image statistic database is constructed before cover selection to store prior knowledge of image embedding suitability and empirical security data for different steganography algorithms and payload. The cover selection model searches for suitable covers under user requirement constraints and randomly chooses a suitable cover (or a group of covers). Finally secret information is embedded into the selected cover.

This cover selection scheme has several advantages against other existing cover selection schemes. The first advantage is low computation consuming as it collects prior knowledge once in advance so that no post verification is required. The second advantage is new steganography algorithms can be easily supported by enriching the image statistic database. The third advantage is it has ability to verify whether a new cover image out of the image database can meet given requirement by computing $\Gamma(\mathrm{X})$.

B. *Construction of Image Statistic Database*

Image statistic database contains the prior knowledge of empirical security of various steganography algorithms with different payload at different image embedding suitability. It has to be constructed before cover selection.

In the database, parameter group <steganography algorithm, payload, minimum OOB, group number> is stored for later cover selection. In this paper, edge adaptive LSBM and HUGO are selected as the steganography algorithms. To reduce the database size, the payload is restricted to 0.10, 0.20, 0.40 and 0.60. Steganography empirical security OOB for each image group is obtained from ensemble classification for each and every composition of steganography algorithm and payload. The OOB used is an average of ten times execution of ensemble training and classification.

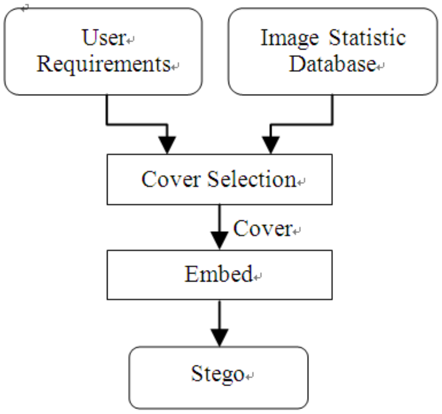

Fig. 2: Proposed Cover Selection Scheme.

In addition to storing the prior knowledge of parameter group <steganography algorithm, payload, minimum OOB, group number>, the database keeps a copy of each and every image's embedding suitability which is ordered in a descending order following embedding suitability. It is used to select suitable covers during cover selection process.

C. *Cover Selection Process*

According to cover selection scheme as shown in Fig. 2, covers are retrieved from image statistic database in the following steps.

User provides constraints of parameter group <steganography algorithm, payload, minimum OOB >;

The scheme first retrieves all the data for the given steganography algorithm, then finds all records whose payload is not less than the given payload, then finds the least group embedding suitability $\overline{\Gamma}(X)$ from where the stored empirical security is the most near to the given minimum OOB;

The scheme selects those images with embedding suitability $\Gamma(X)$ not less than $\overline{\Gamma}(X)$.

In this cover selection scheme, group average embedding suitability $\overline{\Gamma}(X)$ is used as a benchmark to measure the possible least image embedding suitability. Group average embedding suitability can be more accurate if finer image group size is used. Any image with bigger $\Gamma(X)$ compared to $\overline{\Gamma}(X)$are treated as suitable covers meeting user requirement.

The cover selection scheme can also be employed to verify if a new cover image can meet user constraints. Firstly, image embedding suitability $\Gamma(X)$ can be computed following (2) for the given new cover, then find the least group embedding suitability $\overline{\Gamma}(X)$ from image statistic database, finally compare $\Gamma(X)$ against $\overline{\Gamma}(X)$, if the former is not less than the later, the given image cover is a suitable cover to meet user constraints, or else the given cover is not a suitable cover. This kind of verification can be done in very short time and requires little computation power.

D. *Experimental Results*

The image statistic database is constructed with steganography algorithms edge adaptive LSBM and HUGO, payload 0.1, 0.2, 0.4 and 0.6, the empirical security OOB for each composition of <steganography algorithm, payload, image group number>. When computing OOB, the ratio of training versus verification for the ensemble classifier is set to 0.8 as convention.

User requirement constraints of payload and minimum OOB can be found in the experiment result table as row head and column head separately. Minimum OOB that is bigger than 0.40 and payload that is bigger than 0.40, are not listed in the result as not enough covers can be found for classification. Experimental results of edge adaptive LSBM and HUGO are captured in table I and table II separately. In both table I and table II, the column head is the given payload, the row head is the expected OOB, and the table content is the acquired OOB, the double-dashed line "--" means this requirement is not achievable.

It is clearly shown that for both edge adaptive LSBM and HUGO, the acquired empirical security OOB meets the user requirement with only one exception when the user constraint composition is <edge adaptive LSBM, 0.20, 0.40>. This case is very likely due to the image group size 2000 used in image statistic database construction is too large, so the issue could be resolved by using finer group size during image statistic database construction. However, for all other cases, the empirical security is good enough as the acquired OOB exceeds the expectation.

For each steganography algorithm, empirical security OOB descends as the embedding suitability drops. Comparing the result of edge adaptive LSBM to that of HUGO, we can see that the latter performs better than that of the former. This is in line with the security performance of these two steganography algorithms as HUGO performs data embedding in more complex image regions compared to the former. An implication from this comparison is, from application perspective, steganography empirical security can be reached from a composition of steganography algorithm and cover selection.

4. Conclusions

One of the problems of moving steganography from laboratory to real world application is applicable cover selection schemes. This paper tries to resolve this problem by presenting a novel cover selection scheme for spatial image steganography with a new concept of image embedding suitability as the single benchmark to select suitable image covers. This scheme employs prior knowledge of steganalysis to construct an image statistic database and use it to select proper image covers under given requirement constraints of the composition of steganography algorithm, payload and expected minimum empirical security measured in ensemble classifiers' OOB. Extensive experimental results on steganography algorithms of both edge adaptive LSBM and HUGO prove the scheme's effectiveness as the achieved empirical security exceeds the expectation.

TABLE I Result of Edge Adaptive LSBM

Expected OOB	Payload		
	0.10	0.20	0.40
0.40	0.4131	0.3880	--
0.38	0.4007	0.3851	--
0.37	0.3967	0.3746	--
0.36	0.3933	0.3732	--
0.35	0.3921	0.3642	--
0.34	0.3870	0.3525	--
0.33	0.3846	0.3544	--
0.32	0.3822	0.3403	--
0.31	0.3513	0.3384	--
0.30	0.3519	0.3239	--

TABLE Ⅱ Result of HUGO

Expected OOB	Payload		
	0.10	0.20	0.40
0.40	0.4597	0.4278	0.4500
0.38	0.4536	0.4289	0.4469
0.37	0.4517	0.4285	0.3823
0.36	0.4530	0.4279	0.3896
0.35	0.4484	0.4246	0.3915
0.34	0.4462	0.4203	0.3939
0.33	0.4477	0.4215	0.3899
0.32	0.4456	0.4154	0.3832
0.31	0.4211	0.4073	0.3853
0.30	0.4221	0.3998	0.3746

Acknowledgment

Here and now, I would like to extend my sincere thanks to all those who have helped me make this paper possible and better. I am deeply grateful to my honorable master tutor, Fang Yuan, who has checked through my paper with patience and given me instructive suggestions, and he also played an important role in indicating a bright road in my future writing. Then thanks to my beloved family for their loving consideration and support during this study. Finally I am very grateful to my lovely friends who have discussed with me about my paper.

References

[1] T. Filler, J. Judas, and J. Fridrich. "Minimizing additive distortion in steganography using syndrome-trellis codes," *IEEE TIFS*, 6(3), pp.920-935, September 2011.

[2] T. Filler and J. Fridrich. "Gibbs construction in steganography," *IEEE TIFS*, 5(4), pp.705-720, 2010.

[3] W. Luo, F. Huang, and J. Huang. "Edge adaptive image steganography based on LSB matching revisited," *IEEE TIFS*, 5(2):201-214, June 2010.

[4] T. Filler, T. Pevný, and P. Bas. BOSS (Break Our Steganography System). http://www.agents.cz/boss, July 2010.

[5] T. Filler and J. Fridrich. "Design of adaptive steganographic schemes for digital images," In *Proc. SPIE, Elec. Img., Media Watermarking, Sec. and Forensics of Multimedia XIII*, volume 7880, pp.1-14, 2011.

[6] V. Holub, J. Fridrich. "Designing steganographic distortion using directional filters," *IEEE WIFS*, 234-239, 2012.

[7] A. D. Ker, P. Bas, R. Boehme etc. "Moving Steganography and Steganalysis from the Laboratory into the Real World," *1st IH&MMSec. Workshop*, pp.45-58, 2013.

[8] M. Kharrazi, H. Sencar, N. Memon, "Cover Selection for steganographic Embedding," *Proc. of ICIP*, pp.117-121, 2006.

[9] T. Pevný, P. Bas, and J. Fridrich. "Steganalysis by subtractive pixel adjacency matrix," *Proc. of 11th ACM Multimedia & SecurityWorkshop*, pp.75-84, 2009.

[10] J. Kodovský, J. Fridrich, and V. Holub. "Ensemble classifiers for steganalysis of digital media," *IEEE TIFS*, pp.432-444, 2012.